보건학석사 학위논문

Estimating Genetic Influences and Identifying the Genetic Susceptibility Loci For Uterine Cervix HPV infection Status In Korean Women

자궁 경부 부위의 인유두종바이러스 감염 상태와 관련된
유전적 영향 및 감수성 유전자 좌 분석

2015년 2월

서울대학교 보건대학원

보건학과 유전체역학전공

정 다 은

Estimating Genetic Influences and Identifying the Genetic Susceptibility Loci For Uterine Cervix HPV infection Status In Korean Women

지도교수 성 주 헌

이 논문을 정다은 학위논문으로 제출함

2014년 10 월

서울대학교 보건대학원

보건학과 유전체역학전공

정 다 은

정다은의 석사학위논문을 인준함

2014년 12 월

위 원 장 조성일 (인)

부 위 원 장 원성호 (인)

위 원 성주헌 (인)

Abstract

Background: Cervical cancer is the most common cancer in women in some Africa and Asia and the third most common cancer in women worldwide. Human papillomavirus (HPV) infection is the leading cause of cervical lesions and cancers and has a crucial role in the progression of cervical cancers. The burden of HPV infection itself is very substantial in women worldwide. It was reported that approximately 610,000 new patients came out every year worldwide in 2012 and the prevalence of HPV infection was 10.4% in 863 sexually active females aged ranging 20-74 years old in Korea. HPV infection could be heritable because the patterns of progression of HPV infection are different depending on host' genetic susceptibility and environmental factors. Most infections are cured, but 10-20% of women have persistent infections and they could become cervical intraepithelial neoplasm (CINs). 30-40% of CINs are progressed to invasive cancer of the cervix (ICC).

Objective: The aims of this study were to estimate genetic influences and to identify the genetic susceptibility loci with the latest and optimal method to analyze dichotomous trait based on the fact that there is a significant genetic influence on HPV infection status for uterine cervix HPV infection status in Korean Women.

Methods: The 910 study participants aged 25-79 were recruited from the females in the Healthy Twin Study. Prevalence of HPV infected host and odds ratios of HPV infection status by potential risk factors were estimated with a univariate logistic regression analysis. To estimate genetic influences, Intra-class correlation (ICC)

and SOLAR were used and to identify the genetic susceptibility loci, GWAF was used, which is the latest and appropriate method to analyze dichotomous trait.

Results: Among epidemiological factors, only oral contraceptive use was significantly associated with HPV infection status. The ICC coefficient of monozygotic (MZ) twin pairs were higher than mother and daughter pairs and dizygotic (DZ) twin and sibling pairs. Genetic influences also exist moderately (h^2=0.34) in AE model of HPV infection status. In addition, Genetic susceptibility loci of HPV infection status were rs13386094, rs13393102, rs10187756, rs13006868, rs28479291, rs12694475, and rs10165222 in additive model and rs7595290 in general model. They were significantly associated with the phenotype with p-value, $5.4E^{-7}$ and $8.0E^{-8}$ respectively. After analysis, several visualization methods were used to explain the results.

Discussion: Genetic influences of HPV infection status are not that high but significantly heritable. Among the genes from the results in additive model, BROX, CNTN5, GRM8, HMCN1, NEB, PDE4B, PRKAG2, PTPRJ, RHEB, STRC, VNN1, ZBTB7C, ZC3H15, ZNF680, DPP6, RAD51B, and WWOX were statistically associated with diseases such as endometrial cancer and uterine tumor and function of binding of ovarian cancer cell lines (Table 5). In addition, ADAM12, CNTN4, DPP6, GRM8, LRP1B, OBSCN, PDE4D, PLEKHH2, PRKAG2, PTPRJ, RHEB, STRC, STXBP6, VNN1, ZBTB7C, ZC3H15, ZNF680, PDE4D, WWOX, LEPR, and DTD1 were significantly related with the diseases such as endometrial cancer, uterine tumor and the functions in G1/S phase transition of cervical cancer cell lines, early or advanced stage endometriosis, and apoptosis of endometrial cancer cell lines.

Keyword: HPV infection, Heritability, Genetic influences, GWAS, Genetic susceptibility loci, The Healthy Twin Study
Student ID: 2013-21872

Contents

List of Tables

List of Figures

I. Introduction

1. HPV is the key Risk factor of cervical cancer

Cervical cancer is the most common cancer in women in some Africa and Asia and the third most common cancer worldwide in women (1, 2). Approximated 529,000 new patients were calculated in 2008. It has been very well known that Human papillomavirus (HPV) infection is the leading cause of cervical lesions and cancers and has a crucial role in the progression of cervical cancers(3-6). Harald zur Hausen, 'the Father of HPV Virology', first found a relationship between HPV infection and cervical tumor and won the 2008 Nobel Prize(7).

HPV is absolutely species-specific and a small and circular double-stranded DNA virus. HPV infection in genitalia is very common worldwide, usually transmitted by sexual intercourse. Global burden of HPV infection is substantial. Approximately 610,000 new patients came out every year worldwide in 2012 (8). It has been reported that overall prevalence of HPV infection is 10.4% in 863 sexually active females aged ranging 20-74 years old in Busan, South Korea (9).

2. HPV infection has genetic influence.

It has been established that cervical cancer are heritable and where genetic susceptibility loci for cervical cancer are. The women who

have a mother and/ or sister(s) with the history of cervical cancer had significantly higher risk than those who with the second or more relatives (10-14). In addition, several studies on heritability of cervical cancer suggested that having a first degree relative with cervical cancer increases an individual's risk by one to two fold (15).

HPV infection also could be heritable. However, genetic susceptibility loci of HPV infection were not clearly demonstrated. The patterns of progression of HPV infection are different depending on host' genetic susceptibility and environmental factors. Most infections are cured, but 10-20% of women have persistent infections. It is clear that grade 2/3 cervical intraepithelial neoplasms (CINs), which are ranging from low-grade CIN1 to high-grade CIN3, could be progressed to invasive cancer of the cervix (ICC). 30-40% of CINs are progressed to ICC (3,16). Only parts of lesions infected by high-risk HPV do progress to cervical cancers (17, 18). Many observations suggested that host's genetic predisposition has a role in HPV-induced malignant progression as a Non-HPV risk factor (6). Variety of studies have found that mothers in HPV risk group who have a history of genital warts, abnormal Pap smears, or cervical dysplasia have higher risk of HPV vertical transmission to their new born infants (19). However, the genetic influence of HPV infection susceptibility has yet to be cleared.

Sophia S. Wang et al. scrutinized which SNPs were associated with HPV persistence and progression to CIN3 and cervical cancer based on their hypothesis that host genetic variance might differently affect the risk of progression from infection with oncogenic HPV and some SNPs was significantly associated with the two progresses including SNPS residing in genes, RSP19 and PRDX3 (20, 21). The

research team replicated the result in Nigerian women and had same results (22). B.S. Chagas et al. found SNP rs1800896 in interleukin-10 promoter region and a significantly greater susceptibility to progressive cervical lesions in HPV infected women who use oral contraceptives (18). However, a study in Japan performed GWAS to identify SNPs related with susceptibility to persistent infection in genitalia, and the highest p-value was 10^{-5} (17).

3. Risk factor of HPV

It have been studied about the risk factors for HPV infection in previous researches. Above 65 year aged subjects were the highest group of HPV infection (23). Education level, Marital status, Tobacco use, and sexual behavior such as age at sexual debut (24-27) , the possibility of extramarital affairs, history of sexually transmitted disease, vaginal douching and oral contraceptive use were significantly associated with HPV infection status (23, 25-27). A study among Korean women demonstrated Systemic Lupus Erythematosus is a risk factors for high risk HPV infection (28).

4. Purpose

It already has been identified that the HPV infection status depends on an individual's genetic variation. In this study, the aims were to estimate genetic influences and to identify the genetic susceptibility loci with the latest and optimal method to analyze dichotomous trait

based on the fact that there is a significant genetic influences on HPV infection status for uterine cervix HPV infection status in 910 Korean Women.

II. Method

1. Study population

The study participants were recruited from the Healthy Twin Study. They were aged 25-79 and all females. The recruitment in Healthy Twin Study stared out from the 3 hospitals; Samsung Medical Center, Seoul (n=617), Busan Paik Hospital, Busan (n=225), and Dankook University Hospital, Cheonan (n=68) from 2005 and is still followed up. The women who haven't had history of cervical cancer before taking the Papanicolaou (Pap) smear test at that time and their cervico-vaginal smear samples were collected and completed questionnaires. The information from the questionnaires covering demographics, family history, behavioral factors related with lifestyle (oral contraceptive pill use, smoking, and alcohol consumption), social-economic status (educational level, marital status) and clinical information (blood, urine, cervico-vaginal smear samples and stool). Each participants singed on Informed consent (29).

2. Genotyping

The Genomic DNA of 910 subjects was extracted from blood samples at their medical checkup examination, and it was genotyped by Affymetrix Genome-Wide Human SNP Array 6.0

(Affymetrix, CA USA) and Illumina HumanCore-12 BeadChip Kit (Illumina, USA). WTCCC was used to check quality control (QC). Markers that did not meet criteria were not included in final Markers. The criteria was that duplicated, Mendelian inconsistency in $>$ 3 families, Non-Mendelian inconsistency in $>$ 3 families, Minor Allele Frequency(MAF) $<$ 0.01, Hardy-Weinberg Equilibrium(HWE) $<$ 0.001, Genotype missing rate $>$ 0.05 (30). For phasing, SHAPEIT2 was used. Then, IMPUTE2 was used to perform imputation and the genotype that imputation quality score was over 0.9 only remained. Reference haplotype was 1000 Genome haplotypes phase I integrated variant set release GRCh37/hg19 in Asian population.

As for HPV genotyping, the 910 subjects' HPV samples were genotyped. A liquid-based preparation method was used by following standard procedures (ThinPrep®, Hologic, USA and SurePath™, BD Diagnostics-TriPath, USA). For extracting genomic DNA, The Chemagic viral DNA/RNA kit (Chemagen, Germany) was used. After being conducted by PCR, the sequences of HPV were identified by NCBI BLAST program, and genotyping HPV was completed. If HPV infection was detected by at least one of two primer sets, it would considered positive, and was not done by either primer sets, negative. In the result, there were 12 HPV high-risk genotype groups (16, 18, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66), 5 low-risk groups (32, 42, 43, 54 and 70) and seven risk-undetermined HPV genotype groups (30, 67, 74, 81, 84, 90 and 102). Specific subtypes of HPV have different effects (31, 32). Like many other studies, 16 and 18 were the most common and strongly associated with the risk of cervical cancer (23, 31-33). Among 910, 74 subjects were infected by HPV or progressed to CINs.

3. Statistical analysis

The design of this study is a cross-sectional study, consisting of case and control group. In univariate logistic regression analysis, the generalized estimating equation (GEE) model was applied, in which genetic correlations between families were treated as random effects. Odds ratios with 95% confidence intervals were estimated for each (n=910). Software R v. 3.0.2. was used for GEE analysis.

4. Heritability analysis

To estimate genetic influence of HPV infection study, the coefficients of Intra-class correlation (ICC) and heritability estimates were calculated. HPV infection status with low and high risk group in case and with only high risk group were compared in this analysis.

The ICC is a measure of the degree of resemblance between two elements of a particular cluster. It was calculated between the pairs of Monozygotic (MZ) twin, mother and daughter and Dizygotic (DZ) twin, sibling pairs by using the software, SAS v. 9.4. (SAS Institute Inc., USA). The more similarity of HPV infection status between any pairs, the higher the coefficient of ICC

Estimation of Heritability was conducted by a software package, SOLAR that stands for Sequential Oligogenic Linkage Analysis Routines using a genetic variance-components model. Heritability means the proportion of the phenotypic variance (V_P) in a trait that is attributable to the genetic variation (V_G) between individuals.

Depending on definition of genetic variation, heritability could be divided as Broad-sense heritability and narrow-sense heritability. Broad-sense heritability captures the proportion of phenotypic variation due to genetic values that may include effects due to dominance and epistasis.

$$H^2 = V_G/V_P$$

On the other hand, in narrow-sense (additive genetic) heritability, h2 captures only that proportion of genetic variation that is due to additive genetic values (V_A).

$$h^2 = V_A/V_P$$

Shared environmental (household) effects, C^2 captures that proportion of variance due to shared environmental influences.

$$C^2 = V_c/V_P$$

In this study, AE and ACE models were Akaike information criterion

$$(AIC) = 2k - 2\ln(L)$$

k: the number of estimated parameters

L: maximized likelihood

Thus, the AIC value is defined as the log-likelihood penalized by the number of model parameters. Less AIC value, better fitness of model.

5. GWAS analysis

In the previous study, GWAS of HPV was analyzed by plink and FBAT a few years ago. Using plink and FBAT, However, was not a good method to analyze dichotomous data. It is impossible to use total number of samples. Because only 424 samples were used among the total samples, power decreased. In addition, there was no any evidence of genetic influence between the pairs. For identifying the genetic susceptibility loci for uterine cervix HPV infection status, GWAF were used because those method is currently the optimal ways to analyze dichotomous trait correctly. In addition, they made it possible to use total sample size.

GWAF stands for Genome-Wide Association analyses with Family and is also an R package designed mainly for analyzing GWAF. It carries out association tests between single nucleotide polymorphisms (SNPs) and a dichotomous trait with specified genetic model. Logistic regressions is used in this package to analyze dichotomous traits within pedigree correlation via generalized estimating equations. In addition, a general (7, 34-35).

Two models were used in GWAF analysis which are additive model and general model. In additive model, beta coefficients were calculated per 1 copy increase of coded allele, which had been coded as 0, 1 and 2. In general model, beta coefficient for genotype 1 vs. 0, 2 vs. 0 and 2 vs. 1 were estimated respectively. When allele counts are zero or low, general model may be replaced by dominant model in analysis. Dominant model have beta coefficient for genotype 1 and 2 combined vs. genotype 0. The Chi-square statistic of general model is tested whether at least one of the beta10 and

beta20 is not zero (34).

6. Visualization methods

After GWAS analysis, several methods were used to visualize briefly and interpret the result of GWAS analysis. Quantile-Quantile(Q-Q) plot, Manhattan plots, Regional plots and Linkage disequilibrium plot were drawn. These methods are commonly used after GWAS analysis.

The QQ plot was drawn by using R package, 'qqman'. The Q-Q plot make it possible to check whether a result of GWAS follows the expected distribution or not. In other words, it displays the observed association of all SNPs with P-value obtained by GWAS results on the y-axis versus the expected distribution of p-values on the x-axis under the null hypothesis that there is no association. If a result follows expected uniform distribution exactly, it means SNPs are not associated with phenotype. However, the result with strongly associated SNPs will show deviation at the upper-right end of the plot on the diagonal. QQ plot could be used to find systematic bias. Deviation from the diagonal may indicate that data has systematic bias such as population stratification or strong correlation (36).

The Haploview is web-based program and used to analyze HapMap data, choose tag-SNPs and test for association. The Manhattan plots drawn by the HaploView display p-value (-log10) of SNPs on y-axis by chromosome numbers on x-axis. Genetic regions with highly associated SNPs in linkage disequilibrium shows peaks along the

plot (37).

Linkage disequilibrium (LD) plot is also generated by the Haploview. The LD plot implies the high possibility that statistically associated SNPs are in the same LD block formed by historical recombination hot spots. Red represents ‘strong LD between markers and blue does ‘no LD’ (37).

The Regional association plot is drawn to zoom any interesting region of GWAS results by web-based grogram ‘LocusZoom’. Hg19/1000 Genomes 2010 HapMap Asian population were used in the study (38).

Ingenuity Pathway Analysis (IPA) is a web-based software application to identify the pathway, molecular mechanisms, and gene functions. The IPA was used to study the gene products and their diseases and functions related with SNPs in this result (39).

III. Result

1. Statistical result

Table 1. Shows the prevalence of HPV infected host and estimated odds ratios of HPV infection status by potential risk factors such as age, educational level, marital status, oral contraceptive use, smoking, and alcohol intake with a univariate logistic regression analysis via GEE. However, only oral contraceptive use was significantly associated with HPV infection status and the other factors were not in this study. Therefore, oral contraceptive use and age were adjusted in heritability and GWAS analysis.

Table 1. Prevalence of HPV infected host (n=74) and odds ratios estimated of HPV infection status by univariate logistic regression analysis (n=910)

Factor	Category	HPV infected prevalence (%)	OR (95% CI)	*P*
Age(years)	≤34	9 (1.0)	1	
	35-44	26 (2.9)	1.24(0.45-2.03)	0.5873
	45-54	21 (2.3)	1.24(0.43-2.06)	0.5984
	55-64	10 (1.1)	0.68(-0.25-1.61)	0.4168
	≥65	8 (0.9)	1.40()	0.5067
Education level	<Elementary school	11 (1.2)	1	
	High school	42 (4.6)	2.16(1.48-2.85)	0.027
	> College	21 (2.3)	1.76(1.00-2.51)	0.143
Marital status	Never	5 (0.6)	1	
	Married	56 (6.2)	0.67(-0.30-1.64)	0.4155
	Single, divorced and other	13 (1.4)	0.83(-0.27-1.92)	0.7303
Oral contraceptive use*	Never	51 (5.6)	1	
	Ever used	18 (2.0)	1.68(1.12-2.25)	0.0715
	Current used	5 (0.6)	13.47(12.20-14.74)	0.0001
Smoking	Never smoker	68 (7.5)	1	
	Ever smoker	4 (0.4)	1.93(0.84-3.02)	0.2369
	Never smoker	2 (0.2)	0.38(-1.05-1.81)	0.1886
Alcohol intake	Never smoker	28 (3.1)	1	
	Ever smoker	7 (0.9)	1.30(0.43-2.16)	0.5547
	Never smoker	39 (4.3)	1.37(0.87-1.88)	0.2175

*p-value <0.0001 which is determined by chi-square test.

2. Heritability analysis

Table 2 is the result of ICC and heritability estimation of HPV infection status and high risk HPV infection. While HPV itself isn't heritable, HPV infection status could be heritable. Thus, I was interested in how heritable HPV infection susceptibility is. Genetic influence of HPV was calculated by analyzing intra class correlation and heritability of HPV infection status.

ICC coefficients are the Correlation Coefficients between pairs in this study, pairs mean MZ twin pairs, mother and daughter pairs and DZ twin and sibling pairs. In this study, age and oral conceptive use were adjusted. In the result of ICC calculation, you can see coefficients increase as genetic similarity increae between pairs. For high risk HPV infection, the result is also similar to HPV infection status but slightly higher than HPV infection status. Overall, The ICC of MZ twin pairs were higher than mother and daughter pairs and DZ twin and sibling pairs.

In heritability estimates, A, E and C mean additive genetic factor, shared common environmental factor and shared environmental factor respectively. In addition, H2 and c2 stand for genetic component and shared common environmental component between pairs respectively. There were two kinds of model; AE and ACE model. The shared environmental factor implied household effect in ACE model. For HPV infection status, household effect was not significant. As for high risk HPV infection status, both model were not significant.

Genetic influences exist moderately (h2=0.34) in AE model of HPV

Table 2. Intra-class correlations and heritability estimates of HPV and high risk group HPV infection status (n=910).

	Intra-class Correlation Coefficients[1]			Heritability(CI)			
	MZ twin (142 pairs)	M-D (245 pairs)	DZ+ sibling (192 pairs)	*h2*	*c2*	Model fitness[2]	Variance due to final covariates
HPV infection	0.17	0.11	0.10	0.34* (0.00-0.68)	-	*AE model* (AIC: 494.14)	0.02
				0.21 (-0.84-1.25)	0.10 (-0.62-0.82)	*ACE model* (AIC: 494.06)	
H_HPV infection	0.17	0.12	0.13	0.27 (0.07-0.48)	-	*AE model* (AIC: 345.2)	0.00
				0.00	0.21 (0.06-0.36)	*ACE model* (AIC: 345.98)	

CI, confidence interval; MZ, monozygotic; M-D, mother and daughter ; DZ, dizygotic; SE, standard error; H_HPV, high risk group HPV infection; HPV, human papillomavirus.

[1, 2]: Adjusted for age and oral contraceptive use.

* P<0.05

infection status and HPV infection status was significantly heritable.

3. GWAS result

Based on the result that HPV infection status are heritable, genome-wide association study was conducted to find the genetic susceptibility loci. Table 3, 4 describe the GWAF results of HPV infection status with top SNPs associated with HPV infection status and base pairs, chromosome number and p-value. GWAF identified rs13386094, rs13393102, rs10187756, rs13006868, rs28479291, rs12694475, and rs10165222 in additive model (Table 3) and rs7595290 in general model as the most significant SNPs (Table 4) with the phenotype with p-value $5.4E^{-7}$ and $8.0E^{-8}$ respectively.

Table 3. List of significant SNPs associated with HPV infection status in additive model.

SNP	BP	BETA	SE	A1	A2	MAF	REMARK	PVAL	CHR	GENES
rs13386094	220573591	1.60	0.32	A	G	0.035		5.40E-07	2	
rs10165222	220573146	1.60	0.32	T	C	0.035		5.40E-07	2	
rs13393102	220551326	1.60	0.32	A	G	0.035		5.40E-07	2	
rs10187756	220555202	1.60	0.32	T	A	0.035		5.40E-07	2	**SLC4A3, STK11IP, INHA, CHPF, ASIC4, TMEM198, OBSL1, GMPPA, SPEG, DNPEP, PTPRN, RESP18, MIR4268, MIR3132***
rs13006868	220558099	1.60	0.32	A	G	0.035		5.40E-07	2	
rs28479291	220564943	1.60	0.32	C	G	0.035		5.40E-07	2	
rs12694475	220565558	1.60	0.32	A	G	0.035		5.40E-07	2	
rs13395224	220591842	1.60	0.32	A	G	0.035		5.55E-07	2	
rs12694477	220592325	1.60	0.32	A	G	0.035		5.55E-07	2	
rs34364745	220549123	1.60	0.32	C	A	0.035		5.67E-07	2	
rs7564121	220550509	1.60	0.32	C	A	0.035		5.67E-07	2	
rs7595290	220561751	1.65	0.33	A	G	0.035	logistic reg	7.67E-07	2	
rs3784119	68875060	0.88	0.18	C	G	0.436		1.35E-06	14	RAD51B
rs12617025	44404068	0.99	0.22	A	T	0.167		6.53E-06	2	PPM1B
rs60374651	11489955	0.88	0.20	A	T	0.227		7.26E-06	16	LOC388210
rs6945103	151254576	1.92	0.44	C	T	0.016		1.14E-05	7	PRKAG2
rs35587783	78805928	0.83	0.19	A	G	0.242		1.23E-05	16	WWOX
rs267366	174908483	0.88	0.20	A	G	0.169		1.63E-05	5	SFXN1
rs7240386	45578566	0.82	0.19	T	C	0.277		1.69E-05	18	ZBTB7C
rs12401430	210031064	0.98	0.23	G	A	0.129		1.95E-05	1	DIEXF
rs75777472	66522801	1.27	0.30	T	C	0.086		2.24E-05	15	MEGF11
rs7235794	54807183	0.78	0.18	A	G	0.232		2.61E-05	18	LOC102724717
rs71477912	48187839	1.09	0.26	C	A	0.052		2.63E-05	11	PTPRJ
rs2272996	133015271	0.86	0.21	C	T	0.426		2.67E-05	6	VNN1
rs78249280	26975645	2.49	0.59	G	A	0.010	logistic reg	2.76E-05	9	IFT74
rs62544411	26885773	2.36	0.57	G	C	0.012	logistic reg	3.45E-05	9	CAAP1
rs16910873	26904699	2.36	0.57	T	G	0.012	logistic reg	3.45E-05	9	PLAA
rs2942150	204436520	0.80	0.19	G	A	0.174		3.73E-05	1	PIK3C2B
rs77498752	64008144	1.58	0.39	G	C	0.025		4.38E-05	7	ZNF680
rs150815613	64061785	1.58	0.39	C	T	0.026		4.50E-05	7	LOC100128885

rs2515375	99400574	-0.70	0.17	T	C	0.394		4.90E-05	11	CNTN5
rs118047174	152536057	1.12	0.28	C	T	0.069		5.10E-05	2	NEB
rs1713965	57921536	1.36	0.34	T	C	0.030		5.15E-05	4	IGFBP7
rs10117455	126692599	0.93	0.23	G	A	0.102		5.22E-05	9	DENND1A
rs73256510	21676324	1.23	0.31	A	T	0.053	logistic reg	5.41E-05	4	KCNIP4
rs35093122	27288449	-0.75	0.19	G	C	0.445		5.63E-05	10	ANKRD26
rs2291835	222833538	0.95	0.24	T	C	0.149		5.84E-05	1	MIA3
rs77174914	90094423	1.25	0.31	C	T	0.047	logistic reg	6.04E-05	11	DISC1FP1
rs17164985	762285	0.92	0.23	T	C	0.133		6.15E-05	4	PCGF3
rs8089891	6022584	0.89	0.22	G	A	0.116		6.16E-05	18	L3MBTL4
rs147814821	154385994	1.41	0.35	A	T	0.024		6.65E-05	7	DPP6
rs10198525	187355805	1.69	0.42	A	G	0.021		6.67E-05	2	ZC3H15
rs12686284	6919218	0.97	0.24	C	T	0.141		6.76E-05	9	KDM4C
rs141365434	43912682	1.47	0.37	G	C	0.027		7.51E-05	15	STRC
rs150360531	151215611	1.79	0.45	A	G	0.014		7.56E-05	7	RHEB
rs112399429	222899826	0.89	0.23	A	G	0.154		7.87E-05	1	BROX
rs73207751	787947	0.91	0.23	A	G	0.137		8.26E-05	4	CPLX1
rs1322374	186114325	-0.78	0.20	G	C	0.424		8.40E-05	1	HMCN1
rs58090135	44131706	1.47	0.37	G	A	0.026		8.48E-05	15	WDR76
rs10133149	59449594	0.77	0.20	T	C	0.226		8.57E-05	14	LINC01500
rs140215756	44065643	1.47	0.37	G	C	0.026		8.58E-05	15	ELL3
rs7777222	126889537	1.54	0.39	T	C	0.020		8.59E-05	7	GRM8
rs1518404	58268180	1.42	0.36	G	C	0.038		9.01E-05	2	VRK2
rs7986923	102854467	-0.86	0.22	A	G	0.327		9.39E-05	13	FGF14
rs1529989	85143822	0.64	0.16	C	T	0.370		9.44E-05	8	RALYL
rs17011714	222922168	0.88	0.23	T	C	0.154		9.46E-05	1	FAM177B
rs9326069	66257416	0.94	0.24	G	T	0.108		9.51E-05	1	PDE4B
rs4656441	165244152	0.88	0.23	C	T	0.164		9.79E-05	1	LMX1A

*****The closet Genes detected by regional plot

Table 4. List of significant SNPs associated with HPV infection status in general model.

SNP	BP	BETA10	BETA20	BETA21	SE10	SE20	SE21	A1	A2	MAF	REMARK	PVAL	CHR	GENE
rs7595290	220561751	1.66			0.31			A	G	0.035		7.98E-08	2	
rs10165222	220573146	1.60			0.32			C	G	0.035		5.40E-07	2	
rs10187756	220555202	1.60			0.32			T	C	0.035		5.40E-07	2	
rs12694475	220565558	1.60			0.32			A	G	0.035		5.40E-07	2	**SLC4A3, STK11IP, INHA, CHPF, ASIC4, TMEM198, OBSL1, GMPPA, SPEG, DNPEP, PTPRN, RESP18, MIR4268, MIR3132***
rs13006868	220558099	1.60			0.32			A	G	0.035		5.40E-07	2	
rs13386094	220573591	1.60			0.32			A	G	0.035		5.40E-07	2	
rs13393102	220551326	1.60			0.32			A	G	0.035		5.40E-07	2	
rs28479291	220564943	1.60			0.32			T	A	0.035		5.40E-07	2	
rs12694477	220592325	1.60			0.32			A	G	0.035		5.55E-07	2	
rs13395224	220591842	1.60			0.32			A	G	0.035		5.55E-07	2	
rs34364745	220549123	1.60			0.32			C	A	0.035		5.67E-07	2	
rs7564121	220550509	1.60			0.32			C	A	0.035		5.67E-07	2	
rs10976053	7126175	-0.17	1.37	1.53	0.34	0.35	0.33	T	G	0.420		3.51E-06	9	KDM4C
rs7197903	78614508	1.35	0.13	-1.22	0.29	0.74	0.70	T	C	0.227		7.80E-06	16	WWOX
rs5016447	151255099	1.92			0.44			C	T	0.016		1.14E-05	7	PRKAG2
rs139897525	78237846	-0.13	1.49	1.62	0.30	0.35	0.38	G	C	0.254		1.79E-05	5	ARSB
rs6045573	18713002	-0.55	1.67	2.22	0.39	0.44	0.49	G	A	0.170		2.34E-05	20	DTD1
rs9911268	29094253	0.61	2.38	1.77	0.30	0.55	0.60	T	C	0.133		2.38E-05	17	SUZ12P1
rs1036837	67150897	0.11	2.07	1.96	0.32	0.46	0.48	T	A	0.205		2.38E-05	18	DOK6
rs3798800	133012322	0.58	1.64	1.06	0.39	0.39	0.31	T	G	0.417		2.44E-05	6	VNN1
rs79146542	228523171	-0.54	2.01	2.55	0.39	0.48	0.57	C	A	0.141		2.46E-05	1	OBSCN
rs73256510	21676324	1.30			0.31			T	A	0.053		2.80E-05	4	KCNIP4
rs4476244	45572794	0.42	1.95	1.53	0.32	0.43	0.44	T	C	0.155		3.02E-05	18	ZBTB7C

rs7920391	127769667	-0.61	1.14	1.74	0.32	0.36	0.39	G	A	0.331	3.54E-05	10	ADAM12
rs4561818	2386690	-0.19	2.01	2.20	0.33	0.47	0.52	G	A	0.145	3.81E-05	3	CNTN4
rs11133473	57904663	1.50			0.37			G	A	0.028	3.89E-05	4	IGFBP7
rs144567651	64009459	1.58			0.39			A	G	0.025	4.38E-05	7	ZNF680
rs7108003	124966473	0.09	1.68	1.59	0.32	0.39	0.43	C	T	0.176	4.49E-05	11	TMEM218
rs150815613	64061785	1.58			0.39			T	C	0.026	4.50E-05	7	LOC100128885
rs4235478	59521789	0.02	2.61	2.59	0.38	0.59	0.66	C	T	0.101	4.56E-05	5	PDE4D
rs11633990	66522062	1.29			0.32			T	G	0.090	4.57E-05	15	MEGF11
rs73444956	63657800	1.20			0.29			C	T	0.052	4.65E-05	15	CA12
rs13079108	9980469	-0.62	1.84	2.46	0.33	0.50	0.55	A	G	0.210	5.02E-05	3	CRELD1
rs7980896	94289647	0.24	1.36	1.12	0.38	0.37	0.30	A	G	0.411	5.15E-05	12	CRADD
rs10509708	98069164	-0.01	2.13	2.14	0.38	0.49	0.56	T	C	0.090	5.35E-05	10	DNTT
rs72622319	111285519	0.47	2.56	2.09	0.30	0.58	0.59	G	A	0.159	5.38E-05	3	CD96
rs12035604	66056052	-0.95	1.30	2.25	0.36	0.46	0.51	T	A	0.201	5.59E-05	1	LEPR
rs114026714	141120428	1.77			0.44			T	C	0.008	6.05E-05	2	LRP1B
rs10174938	43873144	0.24	1.49	1.24	0.31	0.36	0.35	T	C	0.343	6.08E-05	2	PLEKHH2
rs2942150	204436520	1.20	0.99	-0.22	0.27	0.58	0.57	C	A	0.174	6.41E-05	1	PIK3C2B
rs11626436	25511680	0.28	2.06	1.78	0.31	0.47	0.49	G	C	0.182	6.51E-05	14	STXBP6
rs147814821	154385994	1.41			0.35			A	G	0.024	6.65E-05	7	DPP6
rs10190003	187357465	1.69			0.42			G	A	0.021	6.67E-05	2	ZC3H15
rs10490924	124214448	-1.35	0.01	1.35	0.33	0.32	0.36	T	C	0.428	7.11E-05	10	ARMS2
rs148991660	86877668	1.26			0.32			T	G	0.050	7.21E-05	11	TMEM135
rs141365434	43912682	1.47			0.37			A	G	0.027	7.51E-05	15	STRC
rs150360531	151215611	1.79			0.45			C	T	0.014	7.56E-05	7	RHEB
rs80200516	48097301	1.18			0.30			T	A	0.040	7.96E-05	11	PTPRJ

rs58090135	44131706	1.47			0.37			T	C	0.026	8.48E-05	15	WDR76
rs140215756	44065643	1.47			0.37			G	A	0.026	8.58E-05	15	ELL3
rs7777222	126889537	1.54			0.39			T	C	0.020	8.59E-05	7	GRM8
rs1518404	58268180	1.42			0.36			G	A	0.038	9.01E-05	2	VRK2
rs78249280	26975645	2.61			0.67			A	T	0.010	9.11E-05	9	IFT74
rs167556	77035346	-0.84	1.22	2.07	0.37	0.42	0.48	C	T	0.208	9.39E-05	1	ST6GALNAC3

*The closet Genes detected by regional plot

4. Visualizations

After GWAS analysis, several visualization methods was used to explain the results.

Figure 1 presents the QQ plot of GWAS result of HPV infection status in additive model. It displays the observed association P-value of all SNPs obtained by GWAS results on the y-axis versus the expected distribution on the x-axis. The upper-right points are deviated from the diagonal, which means there are the several significant associated SNPs in the result. Figure 2 also present the similar result in general model.

Figure 3, 4 shows the Manhattan plots for HPV infection status in additive model and general model respectively. The spot is above the red line (5.0E-07) in general model. The number of chromosome 2 have the highest peak in both models.

Figure 5, 6 display the regional plots that zoom in narrow regions for rs13386094 in additive model and in chromosome 2 and rs7595290 in general model both in chromosome 2. The purple diamond indicate the strongest risk SNPs and the other dots are located within LD of r^2 from 0 to 1.

Figure 7, 8 visualize the LD mapping plots with the 12 significant SNP with p-value below (5.0E-07). Red diamond boxes mean the relationship between the two loci among 12 SNPs with D prime. D prime of all pairs were 1, which means the 12 SNPs are located in the same LD block in both models. The yellow star in figure 8 show the locations of rs13386094 of the general model that was top associated SNPs.

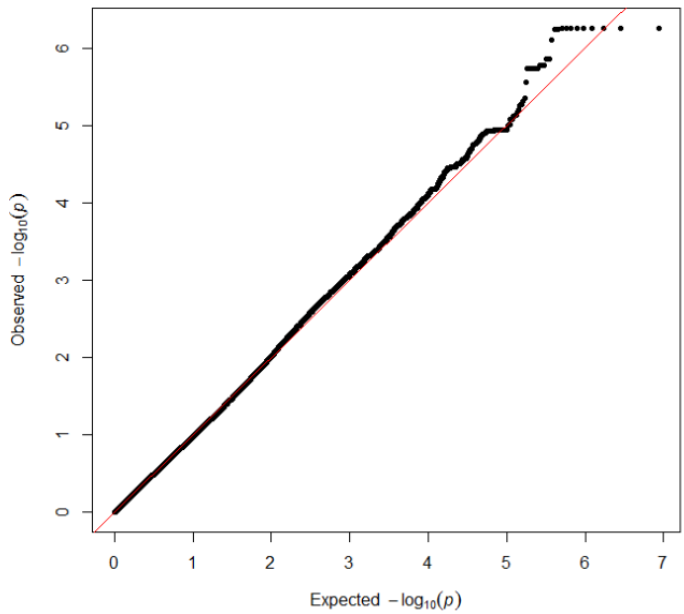

Figure 1. The QQ plot of GWAS result of HPV infection status in additive model.

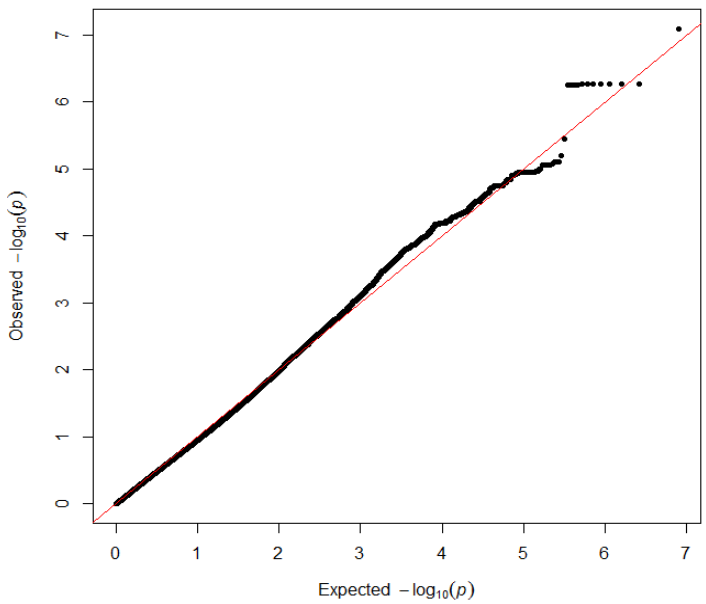

Figure 2. The QQ plot of GWAS result of HPV infection status in general model.

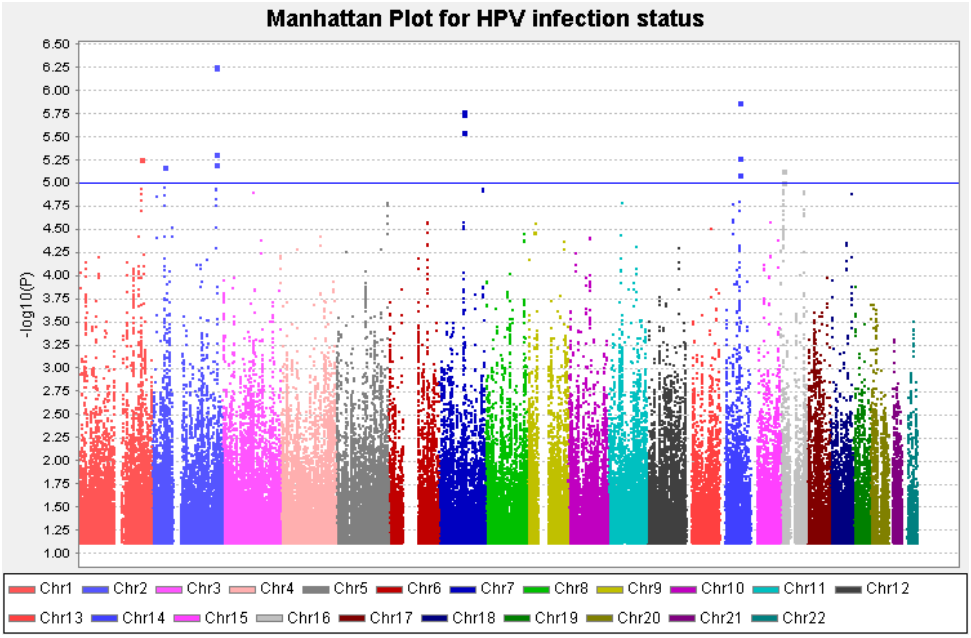

Figure 3. The Manhattan plot for HPV infection status in additive model.

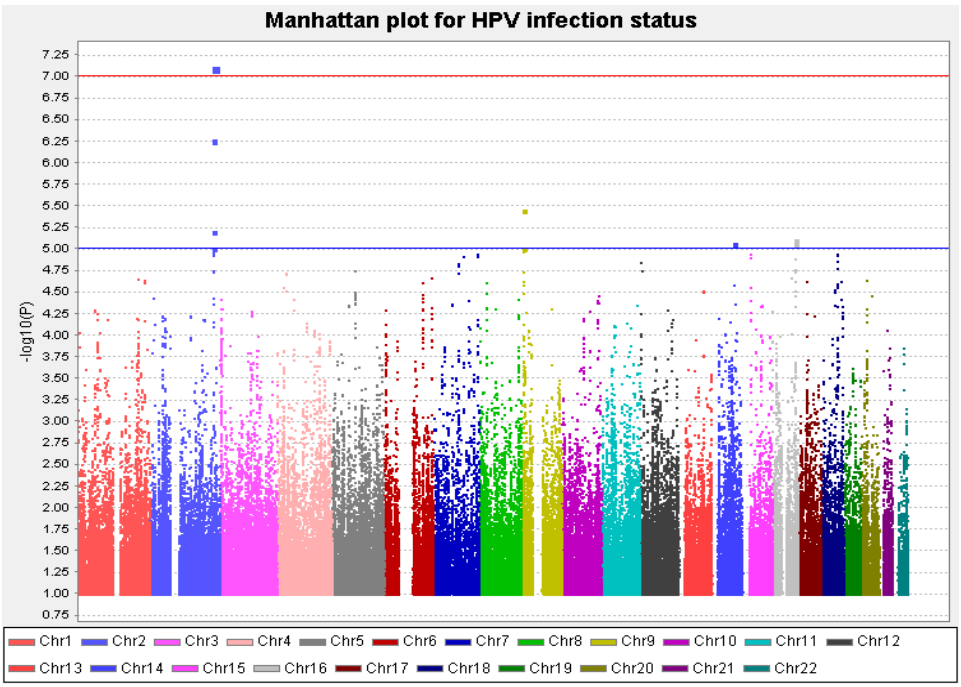

Figure 4. The Manhattan plot for HPV infection status in general model.

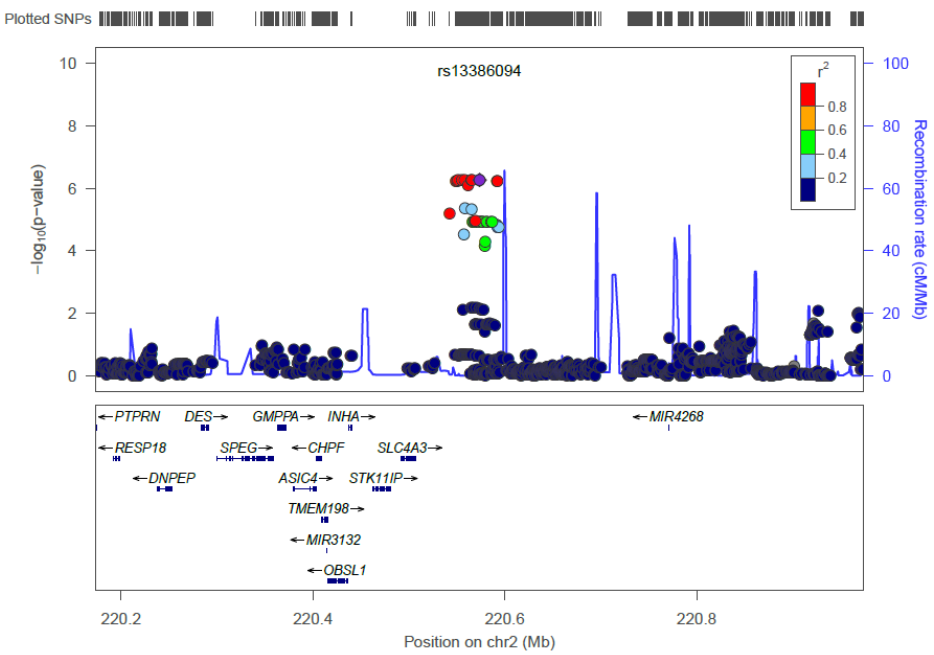

Figure 5. The Regional plot for rs13386094 in chromosome 2 in additive model.

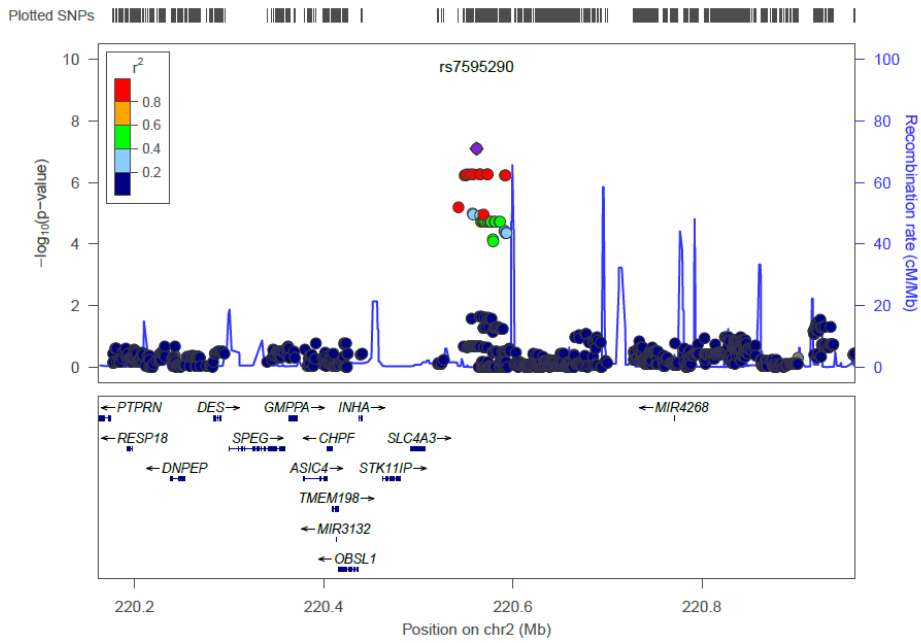

Figure 6. The Regional plot for rs7595290 in chromosome 2 in general model.

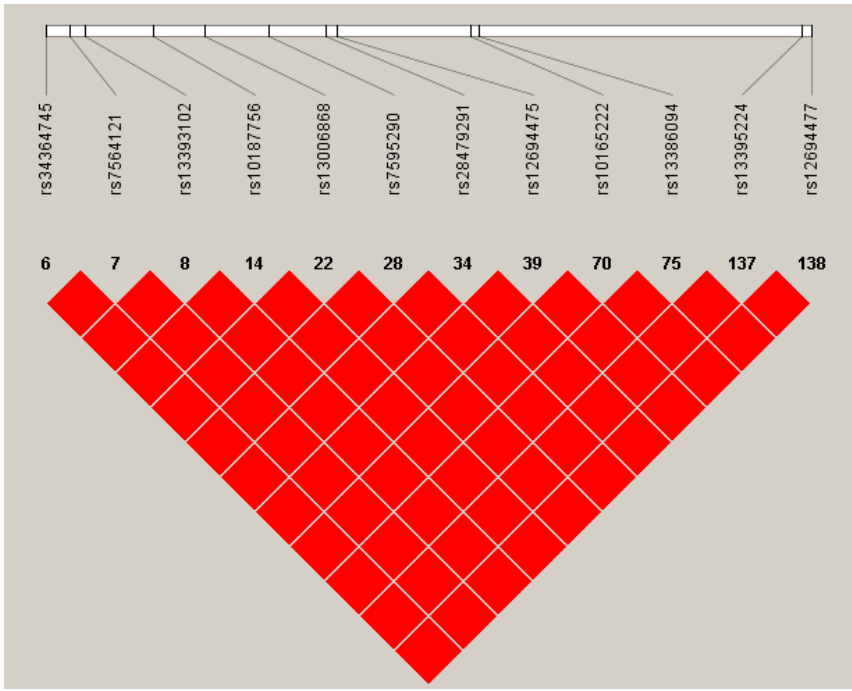

Figure 7. The LD plot in additive model (D'=1)

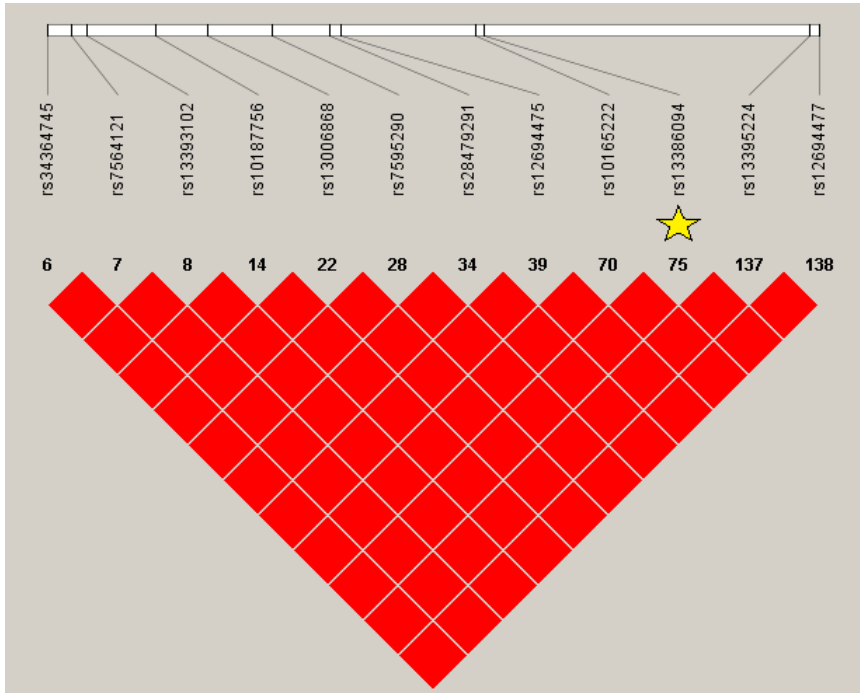

Figure 8. The LD plot in general model (D'=1).

IV. Discussion

The author's finding suggests that genetic influences of HPV infection status in Korean women were not that high but significantly heritable. In further analysis, Genetic susceptibility locus of HPV infection status were rs13393102, rs10187756, rs13006868, rs28479291, rs12694475, rs10165222 and rs13386094 in additive model and rs7595290 in general model as the most significant SNPs. They were significantly associated with the phenotype, 5.40^{-7} and 7.98^{-8}, but they had quite low MAF values and were not included in or closed with meaningful gene regions. Because there is a possibility of false positive in these 12 top SNP, this trait needs to be replicated by other studies. On the other hand, several SNPs were closed to known genes within range of p-value from 1.0^{-6} below 1.0^{-5}.

IPA was used to make these gene lists and their functions which has been disclosed by several studies. Among the genes from the GWAF results in additive model, BROX, CNTN5, GRM8, HMCN1, NEB, PDE4B, PRKAG2, PTPRJ, RHEB, STRC, VNN1, ZBTB7C, ZC3H15, ZNF680, DPP6, RAD51, and WWOX were statistically associated with diseases such as endometrial cancer and uterine tumor and function in binding of ovarian cancer cell lines in Table 5. In addition, ADAM12, CNTN4, DPP6, GRM8, LRP1B, OBSCN, PDE4D, PLEKHH2, PRKAG2, PTPRJ, RHEB, STRC, STXBP6, VNN1, ZBTB7C, ZC3H15, ZNF680, WWOX, LEPR and DTD1 might be related with the diseases such as endometrial cancer and uterine tumor and the functions in G1/S phase transition of cervical cancer cell lines, early or advanced stage endometriosis, binding of ovarian cancer cell lines and apoptosis of endometrial

cancer cell lines in Table 6.

There is only one study about HPV infection, which was reported on the website of National Human Genome Research Institute (http://www.genome.gov/gwastudies). However, it is about head and neck HPV infection not uterine cervix. The Strongest SNP-Risk Allele and its reported gene found in a previous study are in Table 7.

Table 6 also shows that ADAM12, CNTN4, DPP6, IGFBP7, KDM4C, LRP1B, MEGF11, OBSCN, PIK3C2B, STRC, VRK2S which have been reported that they are associated with head and neck cancers. HPV infection is related with 4 % of entire of cancers. Several studies have indicated that infection with specific HPV types is a contributing factor to different types of anogenital cancer, including cervical, anal, penile, and even head, neck and oropharyngeal cancers (3, 40-42). HPV is also known as a main factor of lung, cancer. Iver Petersen et.al concluded that HPV infection is the second contributing risk factor of lung cancer (43).

Table 9 shows the list of measures of LD between two SNPs among the top 12 SNPs. 'L1 and L2' are the two SNP pairs and D' is the measurement of D prime between the two loci. 'LOD' is the log of the likelihood odds ratio, a measure of confidence in the value of D'. 'r2' is the correlation coefficient between the two loci. 'Dist' is the distance (bp) between the loci.

A study indicated that the number of High risk HPV infected women was higher Among HIV positive women than HIV negative. Thus, it suggested that it is helpful for young HIV infected women to make sure to take HPV examination for cervical cancer screening (44). Once young women in Korea start having sexual intercourse HPV

prevalence increase faster than those in the United States and in northern Europe even though they tend to start late (9). It was examined about the relationship between Internet search activity for HPV and HPV vaccine uptake. It is significantly correlated between vaccine search volume and HPV coverage thus, the search term vaccine independently predicted HPV vaccination coverage to support vaccine decision making.(45) Despite of heavy global burden of HPV infection, application of vaccines could reduce significantly the burden of cervical cancer worldwide(6).

There is a limitation in this study. Ideal study design for identifying genetic effect of HPV infection on cervical uterine would be done by considering progression of HPV infection level instead of status. However, it must be attributable if further analysis is conducted based on the result of this cross-sectional study concerning on HPV infection.

To conclude, Heritability, GWAS analysis were performed to estimate genetic influences and to identify the genetic susceptibility loci for uterine cervix HPV infection status in Korean Women. This study identified genetic variants, genes and functions associated with HPV infection status. In addition, it is very meaningful to use the latest and optimal method to analyze dichotomous trait. If this finding is replicated for further analysis with another cohort, it will be more clear the relationship between this trait and the genetic variants found in this study.

Table 5. List of diseases and functions associated with the genes in additive model.

Categories	Diseases or Functions Annotation	p-Value	Genes
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	adenocarcinoma in endometrium	1.80E-02	BROX,CNTN5,GRM8,HMCN1,NEB,PDE4B,PRKAG2,PTPRJ,RHEB,STRC,VNN1,ZBTB7C,ZC3H15,ZNF680
	endometrial cancer	1.57E-02	BROX,CNTN5,DPP6,GRM8,HMCN1,NEB,PDE4B,PRKAG2,PTPRJ,RHEB,STRC,VNN1,ZBTB7C,ZC3H15,ZNF680
	endometrioid carcinoma	1.87E-02	BROX,CNTN5,GRM8,HMCN1,NEB,PDE4B,PRKAG2,PTPRJ,RHEB,STRC,VNN1,ZBTB7C,ZC3H15,ZNF680
	uterine tumor	1.67E-02	BROX,CNTN5,DPP6,GRM8,HMCN1,NEB,PDE4B,PRKAG2,PTPRJ,RAD51B,RHEB,STRC,VNN1,ZBTB7C,ZC3H15,ZNF680
Cell-To-Cell Signaling and Interaction	binding of ovarian cancer cell lines	8.57E-03	WWOX

Table 6. List of diseases and functions associated with the genes in general model.

Categories	Diseases or Functions Annotation	p-Value	Genes
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	adenocarcinoma in endometrium	9.44E-03	CNTN4,GRM8,LRP1B,OBSCN,PDE4D,PLEKHH2,PRKAG2,PTPRJ,RHEB,STRC,STXBP6,VNN1,ZBTB7C,ZC3H15,ZNF680
	endometrial cancer	8.41E-03	CNTN4,DPP6,GRM8,LRP1B,OBSCN,PDE4D,PLEKHH2,PRKAG2,PTPRJ,RHEB,STRC,STXBP6,VNN1,ZBTB7C,ZC3H15,ZNF680
	uterine tumor	9.28E-03	ADAM12,CNTN4,DPP6,GRM8,LRP1B,OBSCN,PDE4D,PLEKHH2,PRKAG2,PTPRJ,RHEB,STRC,STXBP6,VNN1,ZBTB7C,ZC3H15,ZNF680
	endometrioid carcinoma	9.83E-03	CNTN4,GRM8,LRP1B,OBSCN,PDE4D,PLEKHH2,PRKAG2,PTPRJ,RHEB,STRC,STXBP6,VNN1,ZBTB7C,ZC3H15,ZNF680
Cell Cycle	G1/S phase transition of cervical cancer cell lines	1.53E-02	DTD1
Cell Death and Survival	apoptosis of endometrial cancer cell lines	1.75E-02	PDE4D
	apoptosis of ovarian cancer cell lines	1.94E-02	PDE4D,WWOX
Cell-To-Cell Signaling and Interaction	binding of ovarian cancer cell lines	8.79E-03	WWOX
Organismal Injury and Abnormalities, Reproductive System Disease	advanced stage endometriosis	8.79E-03	LEPR
	early stage endometriosis	8.79E-03	LEPR
Head and neck Cancer	head and neck neoplasia	4.64E-02	ADAM12,CNTN4,DPP6,IGFBP7,KDM4C,LRP1B,MEGF11,OBSCN,PIK3C2B,STRC,VRK2

Table 7. List of diseases and functions associated with the closest genes near the top associated SNPs

Categories	Diseases or Functions Annotation	p-Value	Genes
Cell Death and Survival	**cell viability of cervical cancer cell lines**	1.30E-02	PTPRN
	apoptosis of granulosa cells	8.46E-03	INHA
Reproductive System Disease	premature ovarian failure	1.84E-02	INHA
	polycystic ovary syndrome	3.79E-02	INHA

Table 8. List of SNP and genes found in previous study.

First Author	Journal	Study	Region	Chr_pos	Reported Gene(s)	Mapped_gene	Strongest SNP-Risk Allele	p-Value
Chen D	Hum Mol Genet	Genome-wide association study of HPV seropositivity.	6p21.32	32697183	HLA-DQB1	TRNAI25	rs9357152-G	1E-14

Table 9. List of measures of LD between two SNP pairs among the top 12 SNPs.

L1	L2	D'	LOD	r^2	Dist
rs34364745	rs7564121	1 (0.98-1)	148.02	1	1386
rs34364745	rs13393102	1 (0.98-1)	148.02	1	2203
rs34364745	rs10187756	1 (0.98-1)	148.02	1	6079
rs34364745	rs13006868	1 (0.98-1)	148.02	1	8976
rs34364745	rs7595290	1 (0.97-1)	144.17	0.977	12628
rs34364745	rs28479291	1 (0.98-1)	148.02	1	15820
rs34364745	rs12694475	1 (0.98-1)	148.02	1	16435
rs34364745	rs10165222	1 (0.98-1)	148.02	1	24023
rs34364745	rs13386094	1 (0.98-1)	148.02	1	24468
rs34364745	rs13395224	1 (0.98-1)	147.99	1	42719
rs34364745	rs12694477	1 (0.98-1)	147.99	1	43202
rs7564121	rs13393102	1 (0.98-1)	148.02	1	817
rs7564121	rs10187756	1 (0.98-1)	148.02	1	4693

rs7564121	rs13006868	1 (0.98-1)	148.02	1	7590
rs7564121	rs7595290	1 (0.97-1)	144.17	0.977	11242
rs7564121	rs28479291	1 (0.98-1)	148.02	1	14434
rs7564121	rs12694475	1 (0.98-1)	148.02	1	15049
rs7564121	rs10165222	1 (0.98-1)	148.02	1	22637
rs7564121	rs13386094	1 (0.98-1)	148.02	1	23082
rs7564121	rs13395224	1 (0.98-1)	147.99	1	41333
rs7564121	rs12694477	1 (0.98-1)	147.99	1	41816
rs13393102	rs10187756	1 (0.98-1)	148.02	1	3876
rs13393102	rs13006868	1 (0.98-1)	148.02	1	6773
rs13393102	rs7595290	1 (0.97-1)	144.17	0.977	10425
rs13393102	rs28479291	1 (0.98-1)	148.02	1	13617
rs13393102	rs12694475	1 (0.98-1)	148.02	1	14232
rs13393102	rs10165222	1 (0.98-1)	148.02	1	21820
rs13393102	rs13386094	1 (0.98-1)	148.02	1	22265
rs13393102	rs13395224	1 (0.98-1)	147.99	1	40516
rs13393102	rs12694477	1 (0.98-1)	147.99	1	40999
rs10187756	rs13006868	1 (0.98-1)	148.02	1	2897
rs10187756	rs7595290	1 (0.97-1)	144.17	0.977	6549
rs10187756	rs28479291	1 (0.98-1)	148.02	1	9741
rs10187756	rs12694475	1 (0.98-1)	148.02	1	10356
rs10187756	rs10165222	1 (0.98-1)	148.02	1	17944
rs10187756	rs13386094	1 (0.98-1)	148.02	1	18389
rs10187756	rs13395224	1 (0.98-1)	147.99	1	36640
rs10187756	rs12694477	1 (0.98-1)	147.99	1	37123
rs13006868	rs7595290	1 (0.97-1)	144.17	0.977	3652
rs13006868	rs28479291	1 (0.98-1)	148.02	1	6844
rs13006868	rs12694475	1 (0.98-1)	148.02	1	7459
rs13006868	rs10165222	1 (0.98-1)	148.02	1	15047
rs13006868	rs13386094	1 (0.98-1)	148.02	1	15492

rs13006868	rs13395224	1 (0.98-1)	147.99	1	33743
rs13006868	rs12694477	1 (0.98-1)	147.99	1	34226
rs7595290	rs28479291	1 (0.97-1)	144.17	0.977	3192
rs7595290	rs12694475	1 (0.97-1)	144.17	0.977	3807
rs7595290	rs10165222	1 (0.97-1)	144.17	0.977	11395
rs7595290	rs13386094	1 (0.97-1)	144.17	0.977	11840
rs7595290	rs13395224	1 (0.97-1)	144.13	0.977	30091
rs7595290	rs12694477	1 (0.97-1)	144.13	0.977	30574
rs28479291	rs12694475	1 (0.97-1)	148.02	1	615
rs28479291	rs10165222	1 (0.98-1)	148.02	1	8203
rs28479291	rs13386094	1 (0.98-1)	148.02	1	8648
rs28479291	rs13395224	1 (0.98-1)	147.99	1	26899
rs28479291	rs12694477	1 (0.98-1)	147.99	1	27382
rs12694475	rs10165222	1 (0.98-1)	148.02	1	7588
rs12694475	rs13386094	1 (0.98-1)	148.02	1	8033
rs12694475	rs13395224	1 (0.98-1)	147.99	1	26284
rs12694475	rs12694477	1 (0.98-1)	147.99	1	26767
rs10165222	rs13386094	1 (0.98-1)	148.02	1	445
rs10165222	rs13395224	1 (0.98-1)	147.99	1	18696
rs10165222	rs12694477	1 (0.98-1)	147.99	1	19179
rs13386094	rs13395224	1 (0.98-1)	147.99	1	18251
rs13386094	rs12694477	1 (0.98-1)	147.99	1	18734
rs13395224	rs12694477	1 (0.98-1)	147.99	1	483

V. Reference

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International journal of cancer Journal international du cancer. 2010;127(12):2893-917.
2. Kim KS, Park SA, Ko KN, Yi S, Cho YJ. Current status of human papillomavirus vaccines. Clinical and experimental vaccine research. 2014;3(2):168-75.
3. Stanley M. Pathology and epidemiology of HPV infection in females. Gynecologic oncology. 2010;117(2 Suppl):S5-10.
4. Bosch FX, Lorincz A, Muñoz N, Meijer CJLM, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol. 2002.
5. AL. JMMWE. HUMAN PAPILLOMAVIRUS IS A NECESSARY CAUSE OF INVASIVE CERVICAL CANCER WORLDWIDE. JOURNAL OF PATHOLOGY. 1999.
6. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nature reviews Cancer. 2002;2(5):342-50.
7. Mammas IN, Sourvinos G, Spandidos DA. The paediatric story of human papillomavirus (Review). Oncology letters. 2014;8(2):502-6.
8. Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. Vaccine. 2012;30 Suppl 5:F12-23.
9. Shin H-R, Franceschi S, Vaccarella S, et.al J-WR. Prevalence and Determinants of Genital Infection with Papillomavirus, in Female and Male University Students in Busan, South Korea. The Journal of Infectious Diseases. 2004.
10. Engelmark MT, Ivansson EL, Magnusson JJ, Gustavsson IM, Beskow AH, Magnusson PK, et al. Identification of susceptibility loci for cervical carcinoma by genome scan of affected sib-pairs. Human molecular genetics. 2006;15(22):3351-60.
11. Shi Y, Li L, Hu Z, Li S, Wang S, Liu J, et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12.

Nature genetics. 2013;45(8):918-22.
12. Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, et al. Genome-wide association study of susceptibility loci for cervical cancer. Journal of the National Cancer Institute. 2013;105(9):624-33.
13. Wang SS, Bratti MC, Rodriguez AC, Herrero R, Burk RD, Porras C, et al. Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. J Infect Dis. 2009;199(1):20-30.
14. Ivansson EL, Gustavsson IM, Magnusson JJ, Steiner LL, Magnusson PK, Erlich HA, et al. Variants of chemokine receptor 2 and interleukin 4 receptor, but not interleukin 10 or Fas ligand, increase risk of cervical cancer. International journal of cancer Journal international du cancer. 2007;121(11):2451-7.
15. de M. Zelmanowicz A, Hildesheim A. Family history of cancer as a risk factor for cervical carcinoma: a review of the literature. Papillomavirus Report. 2004;15(3):113-20.
16. Moscicki AB, Schiffman M, Burchell A, Albero G, Giuliano AR, Goodman MT, et al. Updating the natural history of human papillomavirus and anogenital cancers. Vaccine. 2012;30 Suppl 5:F24-33.
17. Magnusson PKE, Sparén P, Gyllensten UB. Genetic link to cervical tumours. NATURE. 1999.
18. Chagas BS, Gurgel AP, da Cruz HL, Amaral CM, Cardoso MV, Silva Neto Jda C, et al. An interleukin-10 gene polymorphism associated with the development of cervical lesions in women infected with Human Papillomavirus and using oral contraceptives. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2013;19:32-7.
19. Lee SM, Park JS, Norwitz ER, Koo JN, Oh IH, Park JW, et al. Risk of vertical transmission of human papillomavirus throughout pregnancy: a prospective study. PloS one. 2013;8(6):e66368.
20. Safaeian M, Hildesheim A, Gonzalez P, Yu K, Porras C, Li Q, et al. Single nucleotide polymorphisms in the PRDX3 and RPS19 and risk of HPV persistence and cervical precancer/cancer. PloS one. 2012;7(4):e33619.
21. Wang SS, Gonzalez P, Yu K, Porras C, Li Q, Safaeian M, et al. Common

genetic variants and risk for HPV persistence and progression to cervical cancer. PloS one. 2010;5(1):e8667.
22. Famooto A, Almujtaba M, Dareng E, Akarolo-Anthony S, Ogbonna C, Offiong R, et al. RPS19 and TYMS SNPs and Prevalent High Risk Human Papilloma Virus Infection in Nigerian Women. PloS one. 2013;8(6):e66930.
23. Lee H, Lee DH, Song YM, Lee K, Sung J, Ko G. Risk factors associated with human papillomavirus infection status in a Korean cohort. Epidemiology and infection. 2014;142(8):1579-89.
24. Kim K-D, Kim J-J, Kim S-M, No J-H, Kim Y-B. Prevalence and Determinants of High-risk Human Papillomavirus Infection in Women with High Socioeconomic Status in Seoul, Republic of Korea. Asian Pacific Journal of Cancer Prevention. 2012;13(1):269-73.
25. Kim MA, Oh JK, Kim BW, Chay D, Park DC, Kim SM, et al. Prevalence and seroprevalence of low-risk human papillomavirus in Korean women. Journal of Korean medical science. 2012;27(8):922-8.
26. Shin HR, Lee DH, Herrero R, Smith JS, Vaccarella S, Hong SH, et al. Prevalence of human papillomavirus infection in women in Busan, South Korea. International journal of cancer Journal international du cancer. 2003;103(3):413-21.
27. Kim M-A, Oh J-K, Chay DB, et.al DCP. Prevalence and Seroprevalence of High-Risk Human Papillomavirus Infection. OBSTETRICS & GYNECOLOGY. 2010.
28. Lee YH, Choe JY, Park SH, Park YW, Lee SS, Kang YM, et al. Prevalence of human papilloma virus infections and cervical cytological abnormalities among Korean women with systemic lupus erythematosus. Journal of Korean medical science. 2010;25(10):1431-7.
29. Sung J, Cho S, Lee K, Ha M, Choi E, Choi J, Kim H, et al. Healthy Twin: A Twin-Family Study of Korea — Protocols and Current Status. Twin Research and Human Genetics. 2006; 9(6): 844–8
30. Gombojav B, Song YM, Lee K, Yang S, Kho M, Hwang YC, et al. The Healthy Twin Study, Korea updates: resources for omics and genome epidemiology studies. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2013;16(1):241-5.
31. Frazer IH. HPV vaccines and the prevention of cervical cancer. Update

on Cancer Therapeutics. 2008;3(1):43-8.
32. de Sanjosé S, Diaz M, Castellsagué X, Clifford G, Bruni L, Muñoz N, et al. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. The Lancet Infectious Diseases. 2007;7(7):453-9.
33. Thomas DB, Ray RM, Koetsawang A, et.al NK. Human Papillomaviruses and Cervical Cancer in Bangkok. I. Risk Factors for Invasive Cervical Carcinomas with Human Papillomavirus Types 16 and 18 DNA. American Journal of Epidemiology. 2001.
34. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. Bioinformatics. 2010;26(4):580-1.
35. Kung-Yee Liang; Scott L. Zeger. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika. 1986; 73(1):13-22.
36. S. Turner. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014;
37. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. Cold Spring Harbor protocols. 2009;(10):pdb ip71.
38. Randall J. PruimLZ et.al. LocusZoom: regional visualization of genome-wide association
scan results. BIOINFORMATICS. 2010; 26(18): 2336-2337.
39. Andreas Kra¨mer et.al. Causal analysis approaches in Ingenuity Pathway Analysis. BIOINFORMATICS. 2014; 1-8
40. Megan Scudellari, SEX, CANCER, and a VIRUS. NATURE News Features. 2013; 503: 330-332.
41. Maura L. Gillison, Wayne M. Koch, Randolph B. Capone, Michael Spafford,William H, et al. Evidence for a Causal Association Between Human Papillomavirus and a Subset of Head and Neck Cancers. Journal of the National Cancer Institute. 2000; 92(9): 709–20
42. Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. Journal of the National Cancer Institute. 2008;100(6):407-20.
43. Klein F, Amin Kotb WF, Petersen I. Incidence of human papilloma virus

in lung cancer. Lung cancer. 2009;65(1):13-8.
44. Park EK, Cho H, Lee SH, Lee SG, Lee SY, Kim KH, et al. Human papillomavirus prevalence and genotype distribution among HIV-infected women in Korea. Journal of Korean medical science. 2014;29(1):32-7.
45. Kalichman SC, Kegler C. Vaccine-Related Internet Search Activity Predicts H1N1 and HPV Vaccine Coverage: Implications for Vaccine Acceptance. Journal of health communication. 2014:1-7.

국문초록

자궁 경부 부위의 인유두종바이러스 감염 상태와 관련된 유전적 영향 및 감수성 유전자 좌 분석

서울대학교 보건대학원
보건학과 유전체역학전공
정다은

배경: 자궁경부암은 일부 아프리카와 아시아 지역의 여성에게서 첫 번째로, 전세계적으로는 여성들에서 세 번째로 흔하게 발병하는 암이다. 인유두종바이러스 감염은 자궁경부 상처와 암 발병의 주된 원인이며 자궁경부암으로의 진행에 중요한 원인으로 작용한다. 그리고 여성들에서 인유두종바이러스 감염 자체에 대한 전 세계적인 질병부담도 매우 크다. 2012년에는 전 세계적으로 매년 610,000명의 새로운 환자가 발생하며 863명의 20-74세 성인 여성에서 인유두종바이러스 감염의 유병률이 10.4%임이 보고된 바 있다. 인유두종바이러스 감염은 유전이 될 것이라는 가정이 있는데 그 이유는 인유두종바이러스 감염 후의 진행 양상이 숙주의 유전적인 감수성과 환경요인에 따라 다르게 나타나기 때문이다. 대부분 감염 후 자연적으로 치유가 되지만, 10-20%는 감염이 오랫동안 지속이 되다가 cervical intraepithelial neoplasm (CINs)이 된다. CINs의 30-40%는 invasive cancer of the cervix (ICC) 으로 진행된다.

목적: 본 연구의 목적은 한국인 여성에서 인유두종바이러스 감염 상태에 대한 유적적 영향이 존재한다는 사실을 확인하고, 그러한 사실을 근거로

이분화된 trait를 분석하는 여러 방법 중 현재까지 가장 최신의, 최근의 방법을 이용하여 유전적 감수성 좌를 파악하는데 있다.

방법: Healthy Twin Study에서 25-79세 910명의 여성 연구 참여자가 모집되었다. univariate logistic regression analysis를 이용하여 인유두종바이러스에 감염된 숙주의 유병률과 잠재 위험 요인에 따른 감염 상태의 오즈비를 구하였다. 유전적 영향이 있는지를 평가하기 위해 ICC와 SOLAR가 사용되었고 유전적 감수성 좌를 분석하기 위해 GWAF가 사용되었다.

결론: 여러 역학적 요인들 중에서 경구피임약만 인유두종바이러스 감염 상태와 유의미한 연관이 있었고 본 연구에서 수행한 모든 분석에서 나이와 경구피임약 복용 유무가 보정이 되었다. 일란성 쌍둥이쌍의 ICC 값은 엄마-자녀과 이란성 쌍둥이쌍을 포함한 형제-자매 쌍의 값보다 높았다. 또한, AE model에서 0.34 수준으로 유의미한 유전적 영향이 있음을 확인하였다. 이러한 사실을 근거로 수행한 유전적 감수성 좌에 대한 전장 유전체 연관성 분석에서는 additive model과 general model 각각에서 인유두종바이러스 감염 여부와 가장 높은 유의성을 보인 단일 염기 다형성은 rs13393102 와 rs7595290으로 나타났다. 이들의 유의 수준은 각각 $5.4E^{-7}$ 과 $8.0E^{-8}$ 의 값을 보였고, 분석 후 다양한 visualization 방법을 사용하여 해석을 하였다.

논의: 한국인 여성에서의 인유두종바이러스 감염 상태에 대한 유전적 영향은 많이 높지는 않았지만 유의미한 수준에 있었다. 결과를 통해 나타난 유전자자들과 관련된 질병이나 기능들 중, additive model에서는 BROX, CNTN5, GRM8, HMCN1, NEB, PDE4B, PRKAG2, PTPRJ, RHEB,

STRC, VNN1, ZBTB7C, ZC3H15, ZNF680, DPP6, RAD51B and WWOX 이 endometrial cancer, uterine tumor 와 같은 질병과 ovarian cancer cell lines의 결합 기능에서 통계적으로 유의미한 연관성을 보였다. 또한, general model에서는 ADAM12, CNTN4, DPP6, GRM8, LRP1B, OBSCN, PDE4D, PLEKHH2, PRKAG2, PTPRJ, RHEB, STRC, STXBP6, VNN1, ZBTB7C, ZC3H15, ZNF680, PDE4D, WWOX, LEPR 그리고 DTD1 이 endometrial cancer, uterine tumor 질병 그리고 자궁경부암 cell line의 G1/S 기 전이, 초기 혹은 진행된 수준의 endometriosis, ovarian cancer cell lines의 결합 기능 그리고 endometrial cancer cell line의 세포자살과 같은 기능과 유의미한 연관성을 보였다.

Keyword: HPV infection, Heritability, Genetic influences, GWAS, Genetic susceptibility loci, The Healthy Twin Study
Student ID: 2013-21872